An Overview of Effect Sizes and Meta-analysis

KTDRR Research Evidence Training

Ryan Williams, PhD, Principal Researcher
Joshua R. Polanin, PhD, Principal Researcher
American Institutes for Research

Center on
KNOWLEDGE TRANSLATION FOR DISABILITY & REHABILITATION RESEARCH
at American Institutes for Research

Introduction

- Ryan Williams
  - Principal Researcher, American Institutes for Research (AIR)
  - Principal Investigator: Institute of Education Sciences (IES)-funded meta-analysis focusing on heterogeneity in mathematics intervention effects
  - Co-Principal Investigator: IES-funded training grant on meta-analysis
- Josh Polanin
  - Principal Researcher, AIR
  - Principal Investigator: two National Institute of Justice (NIJ)-funded reviews on school violence
  - Co-Principal Investigator: IES training grant on meta-analysis; IES-funded review on college aid
  - Project Director: What Works Clearinghouse Statistics, Website, and Training (SWAT) contract

Overview

- Effect Size
  - What is an effect size, and why is it important in meta-analysis?
  - How to calculate an effect size depending on what the underlying data are
  - Three varieties: standardized mean-difference, odds ratio, correlation
- Meta-analysis
  - Why synthesize effect sizes?
  - Basics of fixed- and random-effects models
  - Assessing heterogeneity among effect sizes
  - Explaining heterogeneity based on units, treatments, outcomes, and settings (UTOS)
- If time allows—questions at the end

What is an effect size, and why is it important in meta-analysis?

- Start with a thought experiment: Imagine you are interested in understanding whether various programs—all using the same basic tenets—have an impact on an outcome domain.
  - Example: *Therapeutic Effects of Horseback Riding Interventions: A Systematic Review and Meta-analysis* (Stergiou, Tzoufi, Ntzani, Varvarousis, Beris, & Ploumis, 2017)
  - "The purpose of this review was to determine whether therapeutic riding and hippotherapy improve *balance*…" (among many other outcomes)
- How *balance* is measured across each included study probably varies—but each measure focuses on the same underlying construct.

What is an effect size, and why is it important in meta-analysis? (Cont.)

- Meta-analysis expresses the results of each study using a quantitative index of effect size.
- Effect sizes are measures of the strength or magnitude of a relationship of interest.
- Effect sizes have the advantage of being comparable (i.e., they estimate the same thing) across all the studies and therefore can be summarized across studies in the meta-analysis.

What effect size should be calculated?

- Effect sizes can be expressed in many different metrics.
  - *d*, *r*, *OR*, *RR*, etc.
    - The decision about which metric to use is based on what the primary authors report.
- Effect sizes can be unstandardized or standardized.
  - Unstandardized = expressed in measurement units (do not have the properties we need!)
  - Standardized = expressed in standardized measurement units
- Effect sizes should be accompanied by their standard errors; these get calculated along with the effect size and will be used in meta-analytic calculations.

The *d* family

- The standardized mean difference.
- Used when we are interested in two-group comparisons using means.
- Groups could be two experimental groups or, in an observational study, two groups of interest, such as boys versus girls.

The *d* family (Cont.)

- Notation:

  Group means: $\bar{X}_{G1}, \bar{X}_{G2}$

  Group sample sizes: n_{G1}, n_{G2}

  Total sample size: $N = n_{G1} + n_{G2}$

  Group standard deviations: s_{G1}, s_{G2}

The *d* family (Cont.)

- Notation:

  Group means: $\bar{X}_{G1}, \bar{X}_{G2}$

  Group sample sizes: n_{G1}, n_{G2}

  Total sample size: $N = n_{G1} + n_{G2}$

  Group standard deviations: s_{G1}, s_{G2}

$$ES_{sm} = \frac{\bar{X}_{G1} - \bar{X}_{G2}}{s_p}$$

Pooled sample standard deviation (i.e., a weighted average standard deviation)

The *d* family (Cont.)

- The regular effect size formula is biased, especially with small samples.
- We can easily apply a correction for this bias:

$$ES'_{sm} = \left[1 - \frac{3}{4N - 9}\right] ES_{sm}$$

- This produces a Hedges' g effect size, or bias-corrected estimate.

The *d* family (Cont.)

- Each effect size needs a measure of precision:

$$SE_{sm} = \sqrt{\frac{n_{G1} + n_{G2}}{n_{G1} n_{G2}} + \frac{{ES'_{sm}}^2}{2(n_{G1} + n_{G2})}}$$

The *d* family (Cont.)

MULTISYSTEMIC TREATMENT 957

Table 1
Group Means, Standard Deviations, and F Values for Outcome Measures

Measure	Multisystemic therapy Pre	Multisystemic therapy Post	Usual services Pre	Usual services Post	*F*	*p*
Ultimate outcomes						
Arrests[a]					3.94	.050
M	—	0.87	—	1.52		
SD	—	1.34	—	1.55		
Incarceration[a] (in weeks)					7.77	.006
M	—	5.8	—	16.2		
SD	—	13.9	—	19.1		
SRD					4.14	.047
M	11.5	2.9	12.9	8.6		
SD	15.7	5.1	14.3	16.5		

Source: Henggeler, S. W., Melton, G. B. & Smith, L. A. (1992). Family preservation sing multisystemic therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting and Clinical Psychology, 60*(6), *953–961.*

The *d* family (Cont.)

- We can take the values reported in Table 1 and compute an effect size and its standard error:

$$s_p = \sqrt{\frac{(43-1)13.9^2 + (41-1)19.1^2}{(43-1)+(41-1)}} = 16.6$$

$$ES_{sm} = \frac{5.8 - 16.2}{16.6} = -.63$$

$$ES'_{sm} = \left[1 - \frac{3}{4(43+41)-9}\right] - .63 = -.62$$

$$SE_{sm} = \sqrt{\frac{43+41}{43 \times 41} + \frac{-.62^2}{2(43+41)}} = .22$$

The *r* family

- The correlation coefficient, or *r* family effects, may be appropriate when …
  - studies have a continuous outcome measure,
  - study designs assess the relation between a quantitative predictor and the outcome (possibly controlling for covariates), or
  - the analysis uses regression (or the general linear model).

The *r* family (Cont.)

- When using the correlation, you (or a computer program) will do these two things:
  1. Translate it into Fisher's z:

$$Z_r = \frac{1}{2} \ln\left(\frac{1 + r}{1 - r}\right)$$

  2. Calculate the standard error of Z:

$$SE_Z = \frac{1}{\sqrt{n - 3}}$$

The *r* family (Cont.)

- Practically—the formulas are important to know, but you probably won't use them directly.

- Instead—if the correlation is of interest, you will need to locate it in the study and locate its sample size.

The odds ratio family(ish)

- Consider a study in which a treatment group (Tx) and a control group (Cx) are compared with respect to the frequency of a binary characteristic among the participants.
- In each group, we will count how many participants satisfy the binary outcome of interest (e.g., passing a test, graduating, being cured of a disease, etc.).
- The odds ratio is one of a few effect sizes that can be calculated in these scenarios (but it is not the only one).

The odds ratio family(ish) (Cont.)

Study Group	Success	Failure	TOTAL
Treatment	5	14	19
Comparison	6	12	18
TOTAL	11	26	37

Odds of treatment "success" $\frac{5}{14} = .36$

Odds of comparison "success" $\frac{6}{12} = .50$

The "odds ratio" is literally just that—a ratio of two odds—in this case, the odds of treatment success divided by the odds of comparison group success.

$$\frac{.36}{.50} = .72$$

The odds of success in treatment are .72 times the odds of success in control.

The odds ratio family(ish) (Cont.)

- When using the correlation, you (or a computer program) will do these two things:
  1. Transform it to the log odds ratio:

$$Log\ odds\ ratio = \ln(OR)$$

  2. Calculate the log odds ratio standard error:

$$SE_{\ln(OR)} = \sqrt{\frac{1}{Cell\ "a"} + \frac{1}{Cell\ "b"} + \frac{1}{Cell\ "c"} + \frac{1}{Cell\ "d"}}$$

	S	F
Tr	a	b
Co	c	d

Meta-analysis Introduction

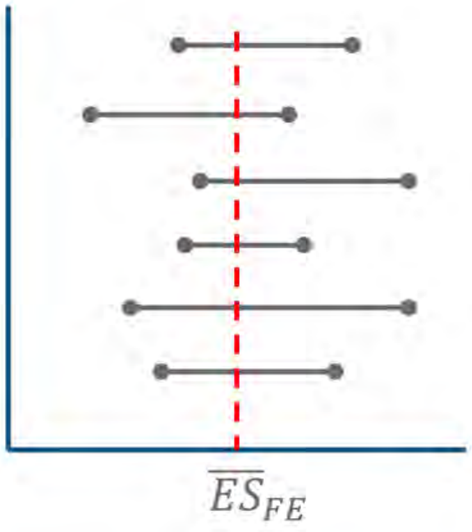
- Computing an average effect is done in two general ways:
  - Fixed-effects model
  - Random-effects model
- Both approaches typically will use an estimate of precision to calculate a weighted mean effect size.
- BUT, the two approaches differ in how they characterize those weights.

Meta-analysis Introduction (Cont.)

- Computing an average effect is done in two general ways:
  - Fixed-effects model
  - Random-effects model
- Both approaches typically will use an estimate of precision to calculate a weighted mean effect size.
- BUT, the two approaches differ in how they characterize those weights.

Meta-analysis Slides

- The fixed-effects model considers ***one*** of variation: sampling variance.



$$\overline{ES}_{FE} = \frac{\sum_{i=1}^{j}(w_i ES_i)}{\sum_{i=1}^{j}(w_i)}$$

$$SE_{FE} = \sqrt{\frac{1}{\sum_{i=1}^{j} w_i}}$$

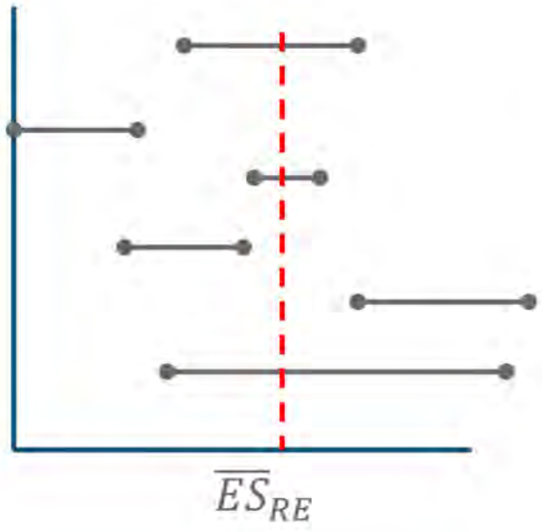
$$w_i = \frac{1}{SE_i^2}$$

Meta-analysis Slides (Cont.)

- This model is helpful when the effect sizes are homogeneous.
- That is, the only reason they are identical is because they have different samples.
- This is a strong assumption.

Meta-analysis Slides (Cont.)

- The random-effects model considers ***two*** sources of variation: sampling variance and between-study variance.



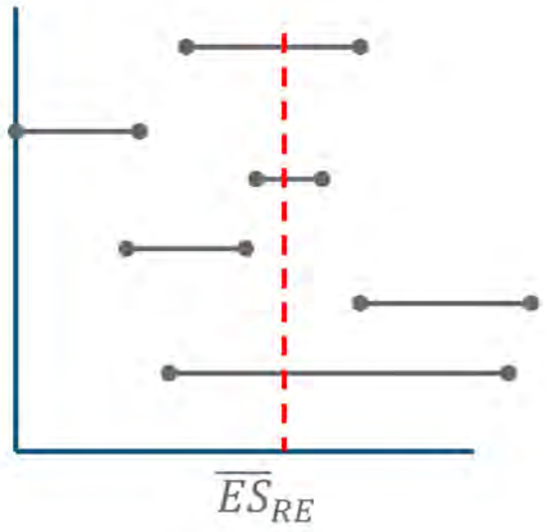
$$\overline{ES}_{RE} = \frac{\Sigma_{i=1}^{j}(w_i ES_i)}{\Sigma_{i=1}^{j}(w_i)}$$

$$SE_{RE} = \sqrt{\frac{1}{\Sigma_{i=1}^{j} w_i}}$$

$$w_i = \frac{1}{SE_i^2 + \hat{\tau}^2}$$

Meta-analysis Slides (Cont.)

- The random-effects model considers ***two*** sources of variation: sampling variance and between-study variance.



$$\overline{ES}_{RE} = \frac{\sum_{i=1}^{j}(w_i ES_i)}{\sum_{i=1}^{j}(w_i)} \qquad SE_{RE} = \sqrt{\frac{1}{\sum_{i=1}^{j} w_i}}$$

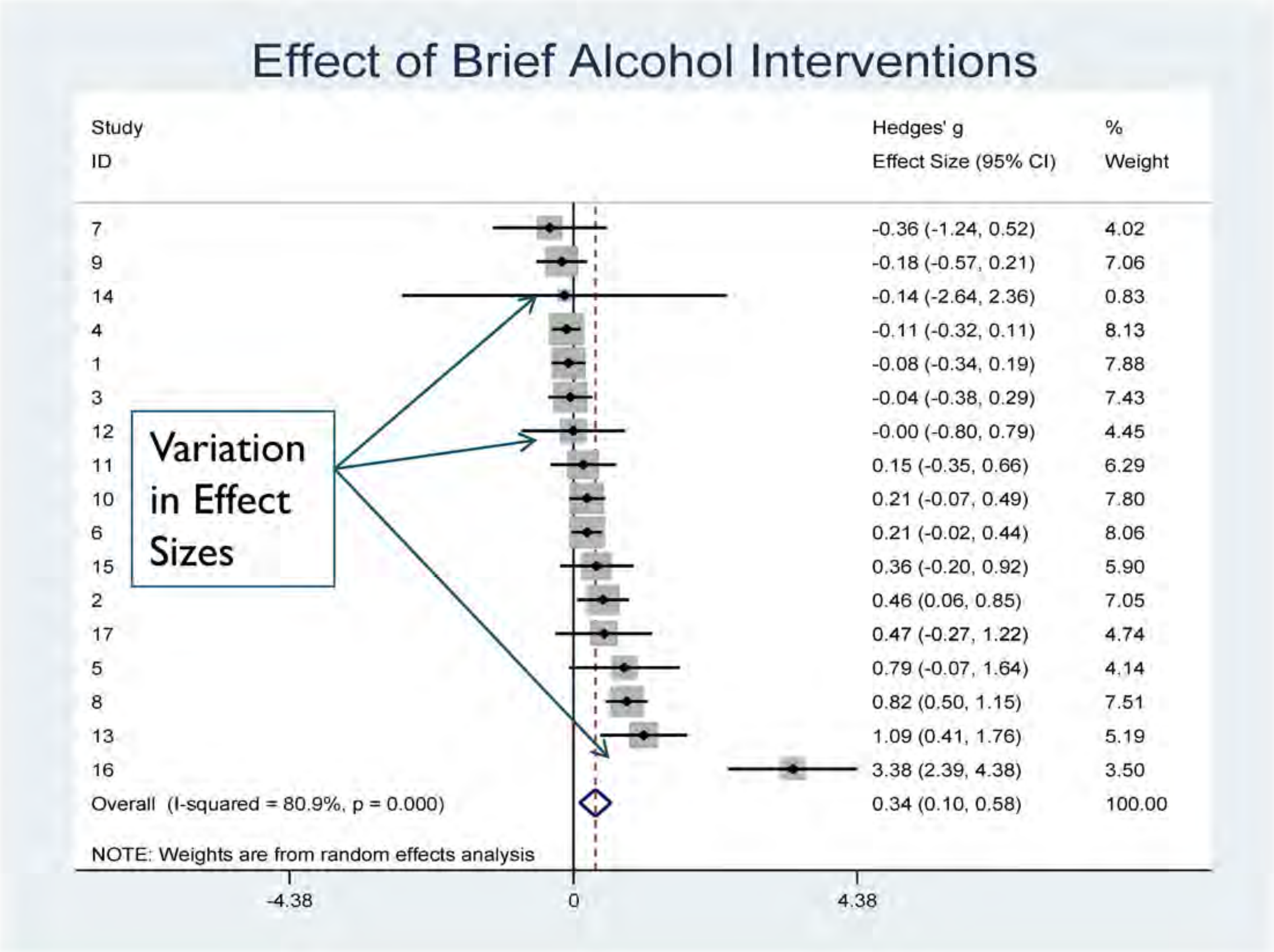
$$w_i = \frac{1}{SE_i^2 + \hat{\tau}^2}$$

Estimated heterogeneity

Meta-analysis Slides (Cont.)

- Because the random-effects model uses two sources of variation, the standard error around the mean effect will be larger than the standard error for an equally sized fixed-effects model.
- BUT, unless the multiple effect sizes are coming from maximum-control conditions (i.e., carefully scripted, laboratory-type replications), it is hard to rule out a random-effects model.
- For this reason, we strongly encourage using a random-effects model. If your data are statistically homogeneous, it won’t hurt you!

What is effect size heterogeneity?



Source: Hennessy & Tanner-Smith, *Prevention Science*, 2015

Quantifying heterogeneity

- Three primary statistics:
  - Q – Tells us if the variation is different from chance.
  - τ^2 – Tells us the magnitude of the variation.
  - I^2 – Tells us the proportion of true variation among effects (taking into account the possibility of random variation).

- These allow us to turn visual information into quantitative information.

Explaining heterogeneity

- What to do with sufficient heterogeneity?

- One-way moderator analyses (old approach)
- One-variable meta-regression (old approach)

- Multiple variable meta-regression (best approach)
  - Control for confounding factors
  - Reduce type 1 errors
  - Easier to interpret

Thank you!

Please take a few minutes to respond to the brief Evaluation Survey:

https://www.surveygizmo.com/s3/4552615/Webcast-Eval-Effect-Sizes-Meta-Analysis

Ryan Williams: rwilliams@air.org

Joshua Polanin: jpolanin@air.org

www.ktdrr.org

ktdrr@air.org

4700 Mueller Blvd, Austin, TX 78723

800.266.1832

Center on
KNOWLEDGE TRANSLATION FOR DISABILITY & REHABILITATION RESEARCH
at American Institutes for Research

The contents of this presentation were developed under grant number 90DPKT0001 from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS). The contents of this presentation do not necessarily represent the policy of NIDILRR, ACL, HHS, and you should not assume endorsement by the Federal Government.